Advanced Network Technologies

Multimedia 2/2

Dr. Wei Bao | Lecturer
School of Computer Science

THE UNIVERSITY OF SYDNEY

Insert Partner Logo
- Delete if not required

SIP: Session Initiation Protocol [RFC 3261]

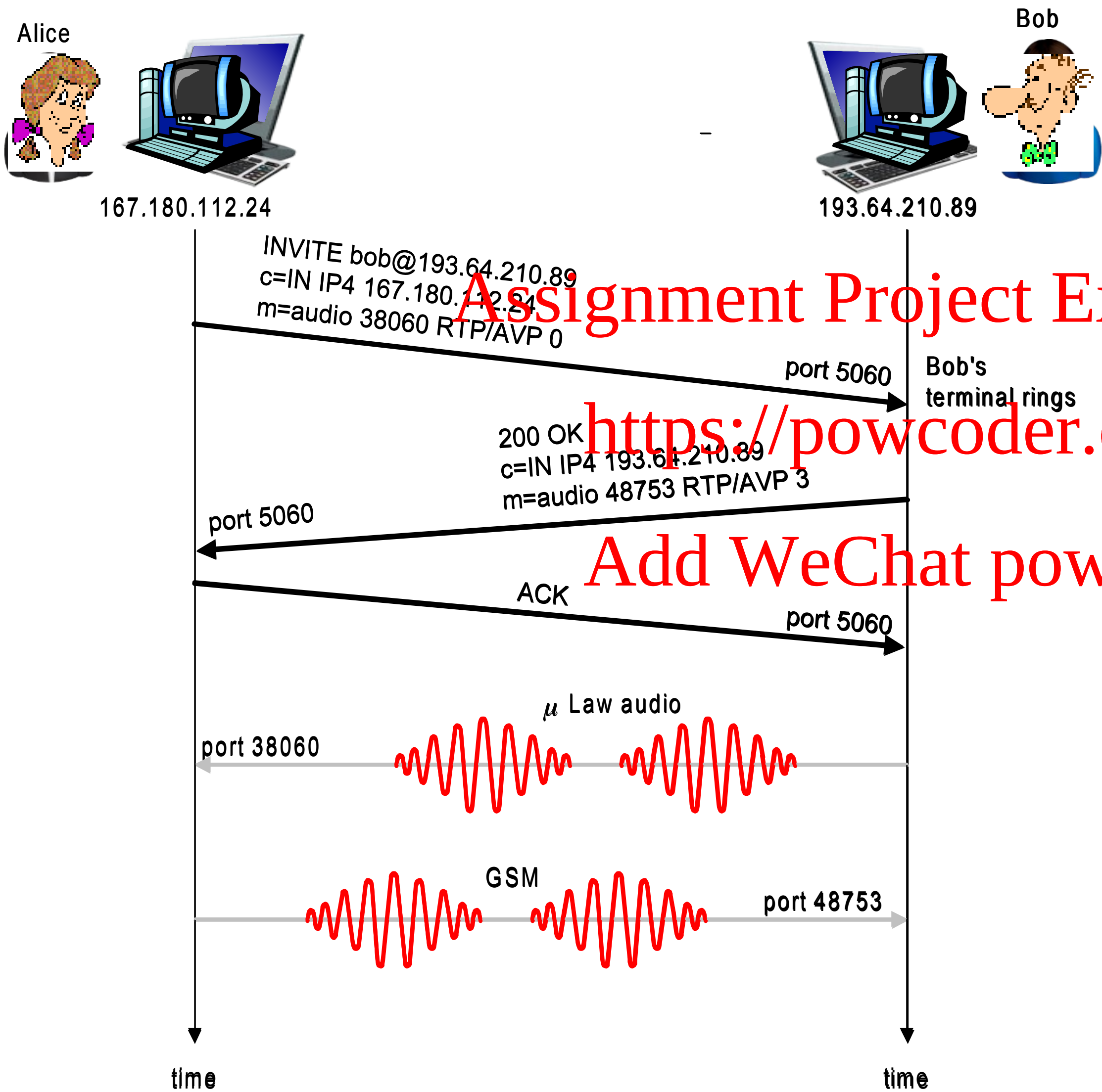
long-term vision:

- all telephone calls, video conference calls take place over Internet
- people identified by names or e-mail addresses, rather than by phone numbers
- can reach callee *(if callee so desires),* no matter where callee roams, no matter what IP device callee is currently using

SIP services

- SIP provides mechanisms for call setup:
  - for caller to let callee know she wants to establish a call
  - so caller, callee can agree on media type, encoding
  - to end call
- determine current IP address of callee:
  - maps mnemonic identifier to current IP address
- call management:
  - add new media streams during call
  - change encoding during call
  - invite others
  - transfer, hold calls

Example: setting up call to known IP address



- Alice's SIP invite message indicates her port number, IP address, encoding she prefers to receive (PCM μlaw)
- Bob's 200 OK message indicates his port number, IP address, preferred encoding (GSM)
- SIP messages can be sent over TCP or UDP; here sent over RTP/UDP
- Default SIP port # is 5060
- Actually, Bob and Alice talks simultaneoulsy
- SIP is out-of-band

Setting up a call (cont'd)

- codec negotiation:
  - suppose Bob doesn't have PCM µlaw encoder
  - Bob will instead reply with 606 Not Acceptable Reply, listing his encoders. Alice can then send new INVITE message, advertising different encoder
- rejecting a call
  - Bob can reject with replies "busy," "gone," "payment required," "forbidden"
- media can be sent over RTP or some other protocol

Name translation, user location

- caller wants to call callee, but only has callee's name or e-mail address.
- need to get IP address of callee's current host:
  - user moves around
  - Dynamic Host Configuration Protocol (DHCP) (dynamically assign IP address)
  - user has different IP devices (PC, smartphone, car device)
- result can be based on:
  - time of day (work, home)
  - caller (don't want boss to call you at home)
  - status of callee (calls sent to voicemail when callee is already talking to someone)

SIP registrar

- one function of SIP server: registrar
- when Bob starts SIP client, client sends SIP REGISTER message to Bob's registrar server

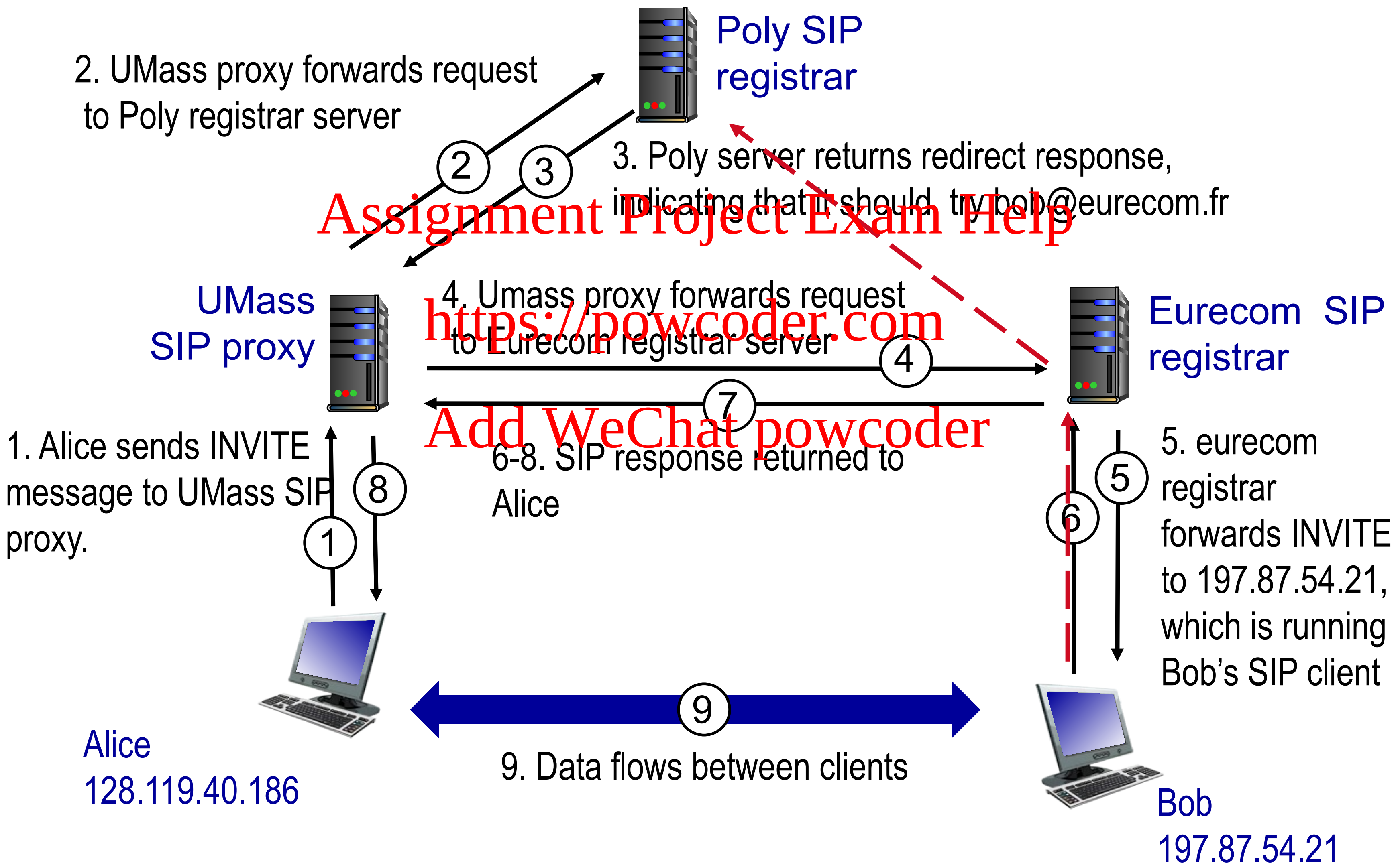
register message:


```
REGISTER sip:domain.com SIP/2.0
Via: SIP/2.0/UDP 193.64.210.89
From: sip:bob@domain.com
To: sip:bob@domain.com
Expires: 3600
```

SIP proxy

- another function of SIP server: *proxy*
- Alice sends invite message to her proxy server
  - contains address sip:bob@domain.com
  - proxy responsible for routing SIP messages to callee, possibly through multiple proxies
- Bob sends response back through same set of SIP proxies
- proxy returns Bob's SIP response message to Alice
  - contains Bob's IP address
- SIP proxy analogous to local DNS server

SIP example: alice@umass.edu calls bob@poly.edu



Network support for Multimedia

Network support for multimedia

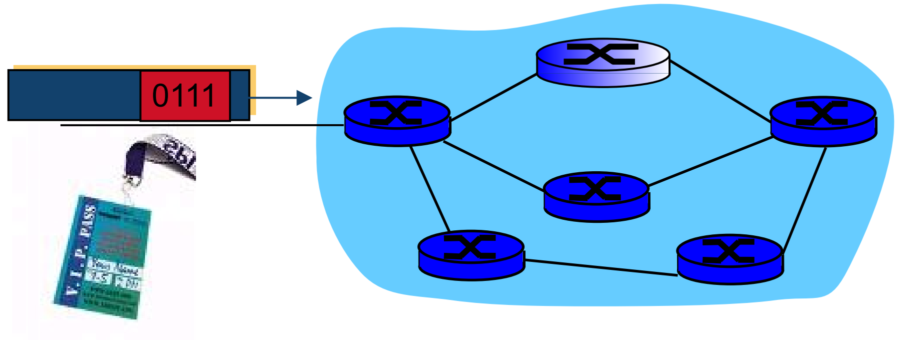
Three broad approaches towards providing network-level support for multimedia apps

Approach	Granularity	Guarantee	Mechanisms	Comple x	Deployed?
1. Making best of best effort service	All traffic treated equally	None	No network support(all at app)	low	everywhere
2. Differentiated service	Traffic class	None or soft	Packet mark, scheduling, policing	medium	some
3. Per-connection QoS	Per-connection flow	Soft or hard after flow admitted	Packet mark, scheduling policing	high	Little to none

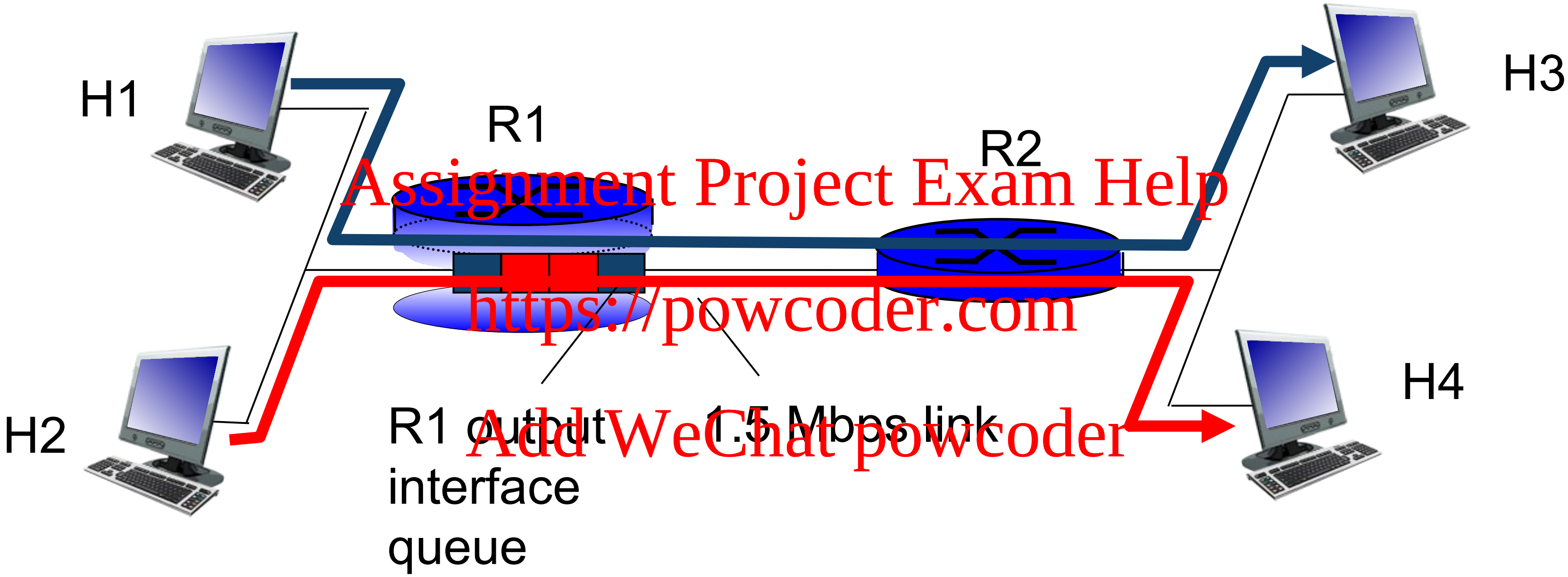
Providing multiple classes of services

- thus far: making the best of best effort service
  - one-size fits all service model
- alternative: multiple classes of service
  - partition traffic into classes
  - network treats different classes of traffic differently (analogy: VIP service versus regular service)

- granularity: differential service among multiple classes, not among individual connections
- How: ToS bits

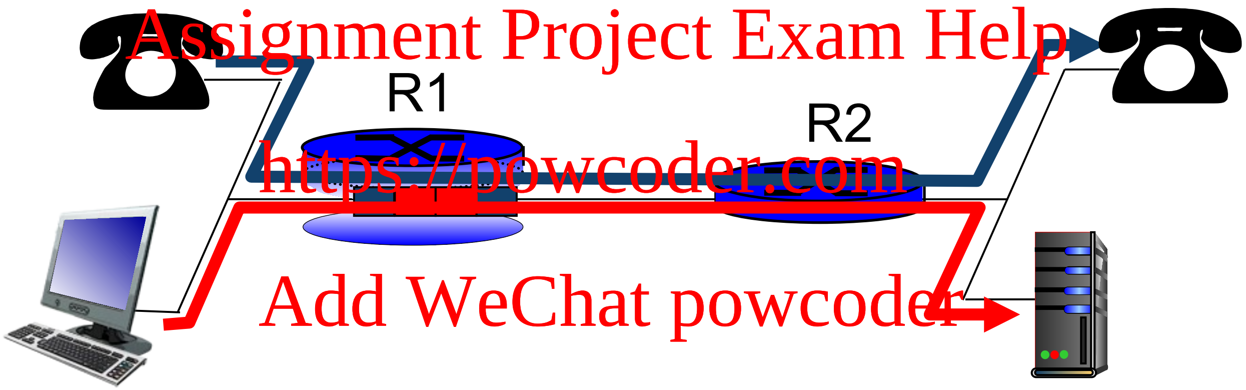


Multiple classes of services: scenarios



Scenario 1: mixed HTTP and VoIP

- example: 1Mbps VoIP (Video and Voice), HTTP share 1.5 Mbps link.
  - HTTP bursts can congest router, cause video/audio loss
  - want to give priority to audio over HTTP

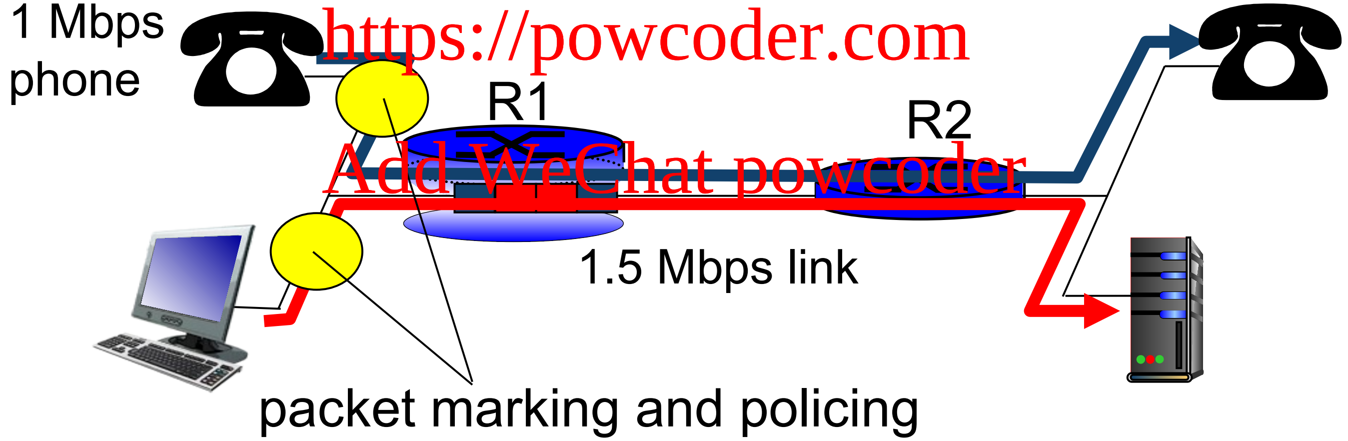


Principle 1

packet marking needed for router to distinguish between different classes; and new router policy to treat packets accordingly

Principles for QOS guarantees

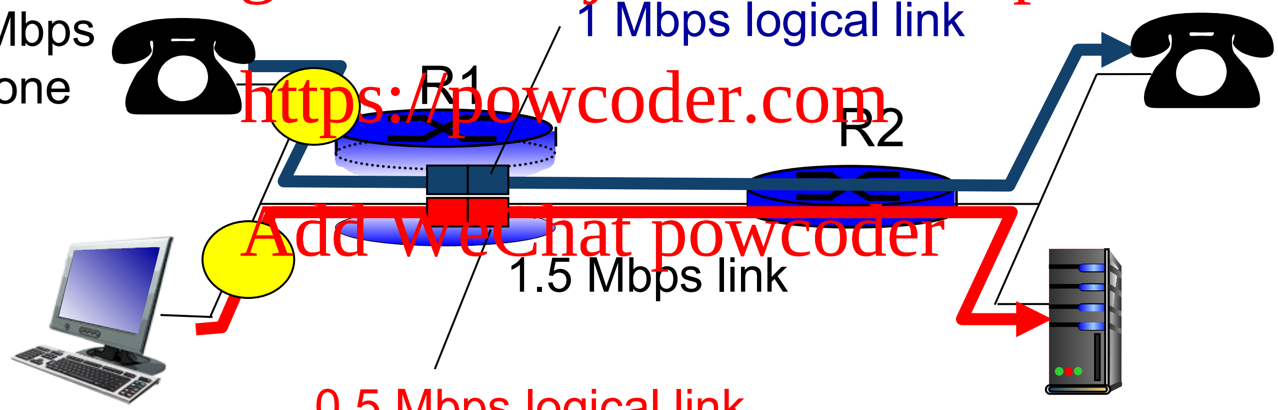
- what if applications misbehave (VoIP sends higher than declared rate)
  - policing: force source adherence to bandwidth allocations
- *marking*, *policing*



Principle 2
provide protection for one class from others

Principles for QOS guarantees (con't)

- allocating *fixed* (non-sharable) bandwidth to flow: *inefficient* use of bandwidth if flows doesn't use its allocation

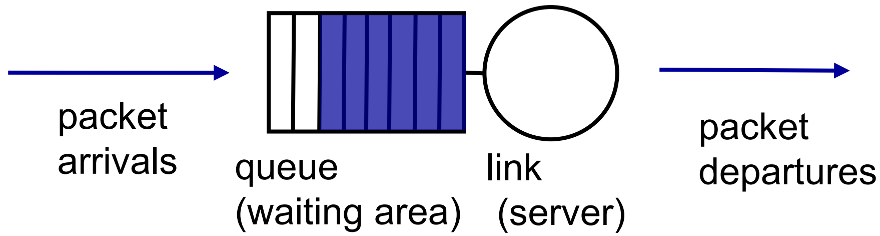


Principle 3

while providing protection, it is desirable to use resources as efficiently as possible

Scheduling and policing mechanisms

- *scheduling:* choose next packet to send on link
- *FIFO (first in first out) scheduling:* send in order of arrival to queue
  - real-world example?
  - *discard policy:* if packet arrives to full queue: who to discard?
    - *tail drop:* drop arriving packet
    - *priority:* drop/remove on priority basis
    - *random:* drop/remove randomly

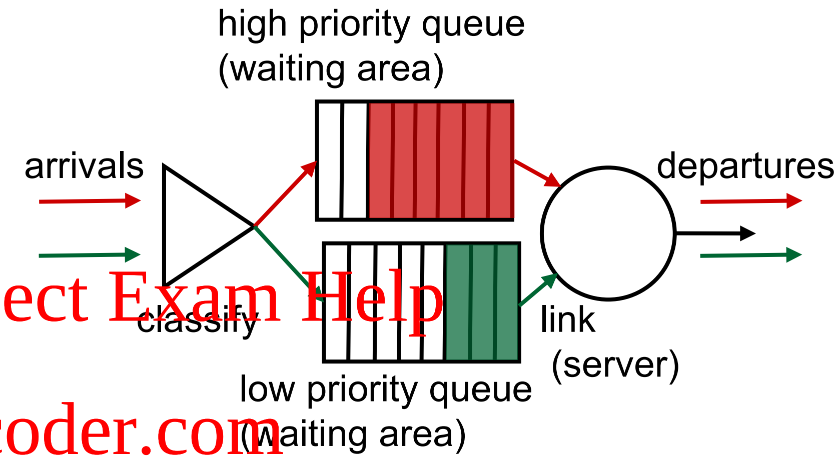


THE UNIVERSITY OF SYDNEY

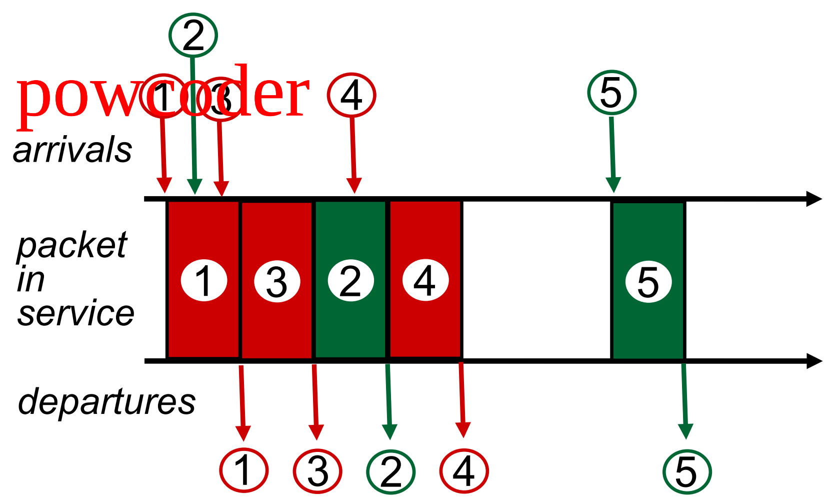
Scheduling policies: priority

priority scheduling: send highest priority queued packet

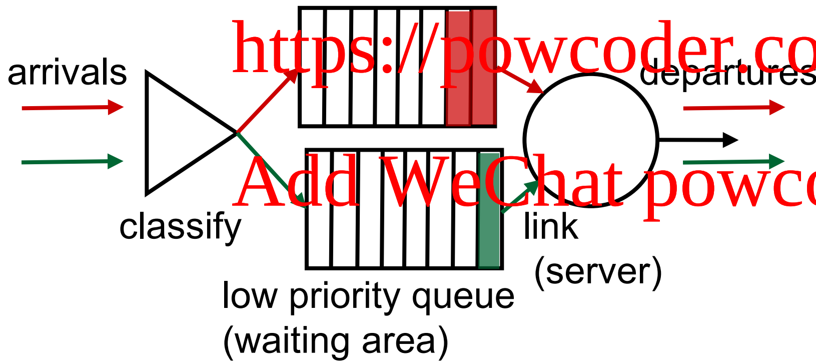
non-preemptive



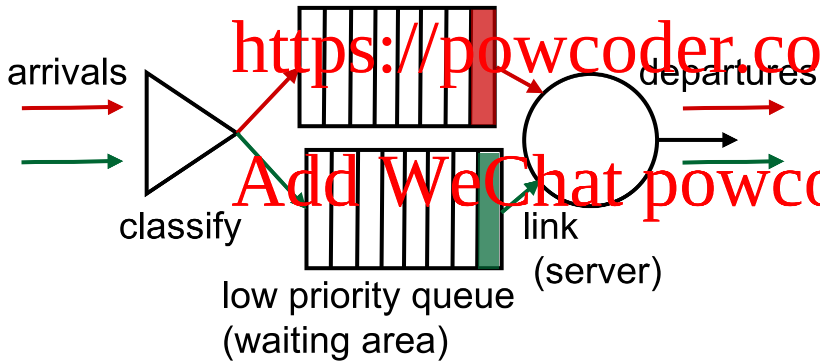
- multiple *classes*, with different priorities
  - class may depend on marking or other header info, e.g. IP source/dest, port numbers, etc.
  - real world example?



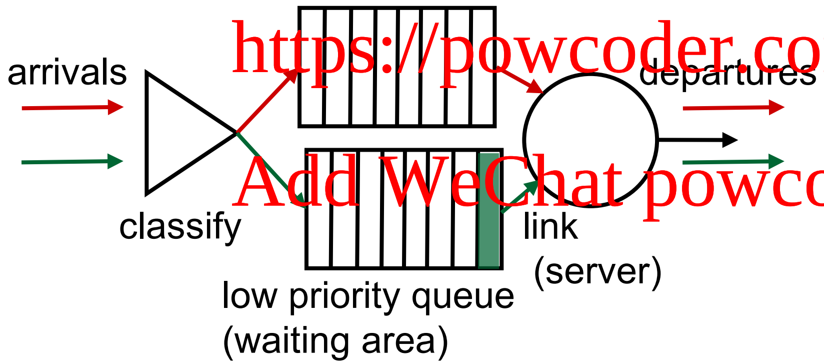
Scheduling policies: priority



Scheduling policies: priority



Scheduling policies: priority

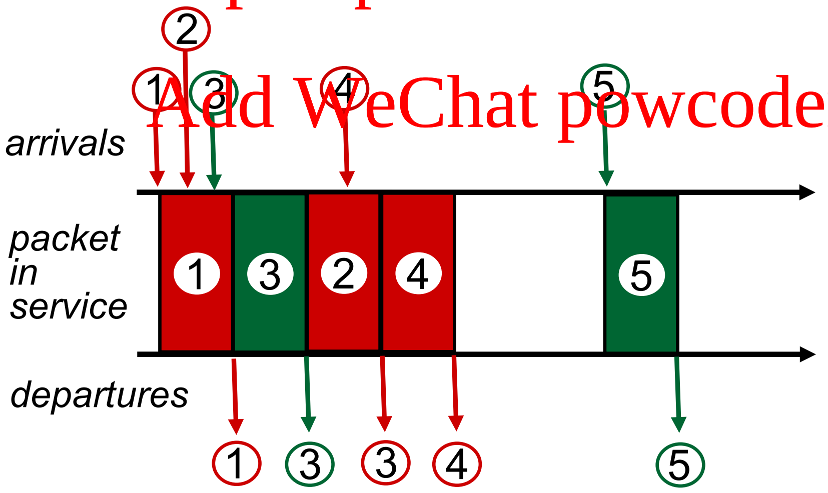


Scheduling policies (con't)

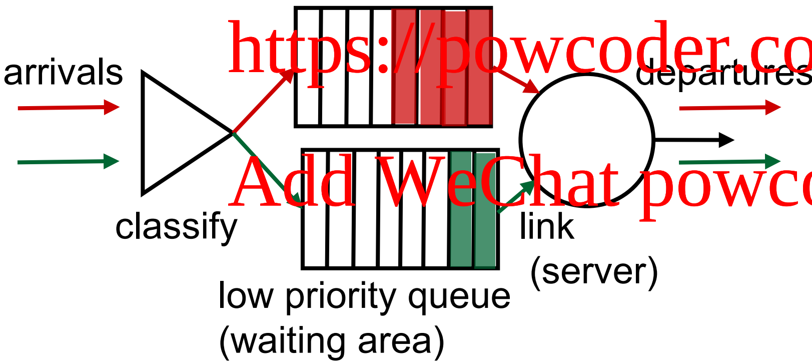
Round Robin (RR) scheduling:

› multiple classes, with equal priority

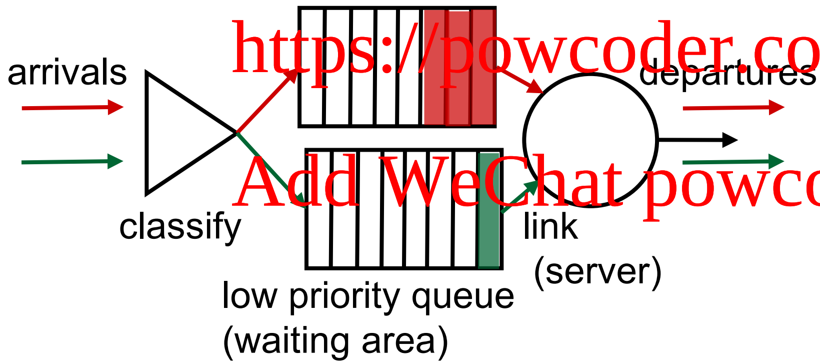
› cyclically scan class queues, sending one complete packet from each class (if available)



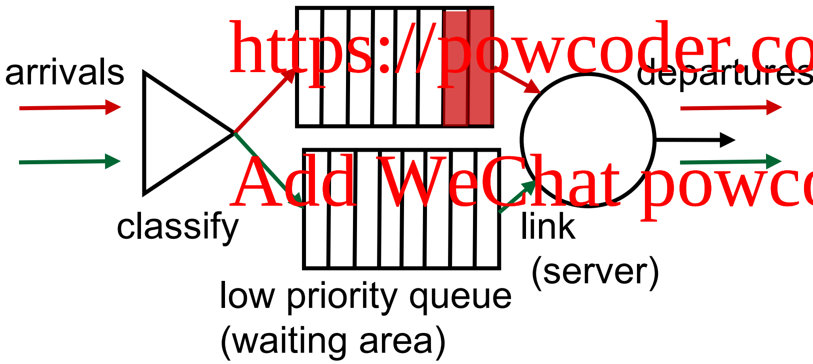
Scheduling policies: RR



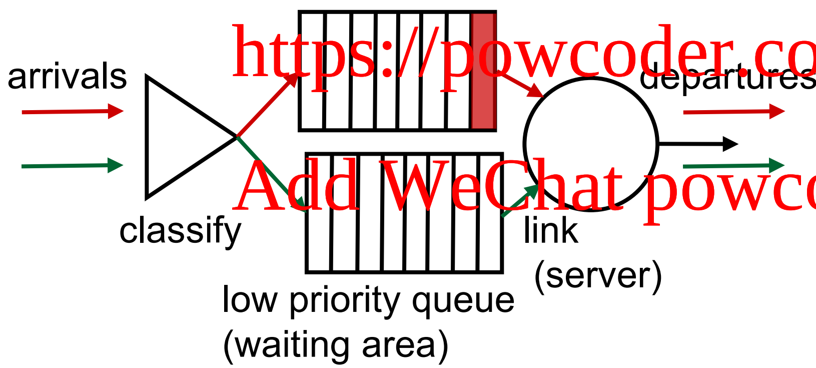
Scheduling policies: RR



Scheduling policies: RR



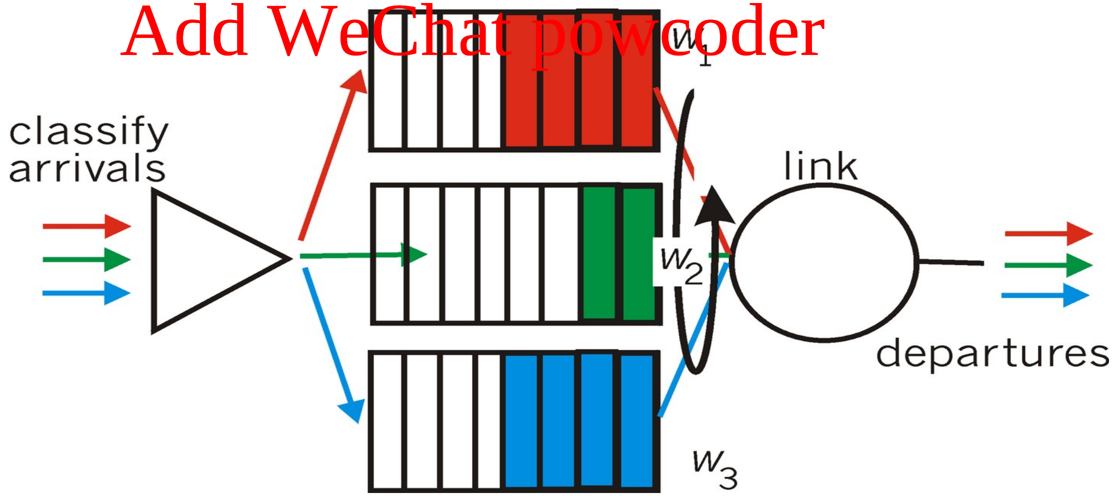
Scheduling policies: RR



Scheduling policies (con't)

Weighted Fair Queuing (WFQ):

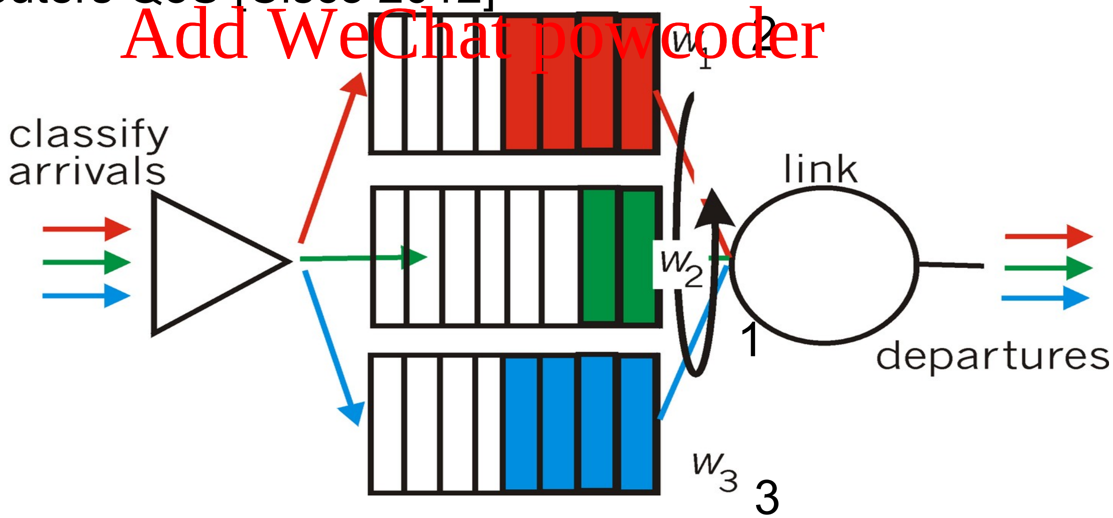
- generalized Round Robin
- each class gets weighted amount of service in each cycle



Scheduling policies (con't)

Weighted Fair Queuing (WFQ):

- Each class *i* is assigned a weight w_i
- *Guarantee*: if there are class *i* packets to send (during some interval) then class *i* receives a fraction of service which is
  $w_i / (\sum w_j)$
- On a link with transmission rate *R*, class *i* achieves throughput
  $Rw_i / (\sum w_j)$
- WFQ is part of routers QoS [Cisco 2012]



Scheduling policies (con't)

Example:

One link has capacity 1 Mbps. Three flows: Flow 1 is ensured with 0.5 Mbps data rate; Flow 2 is ensured with 0.25 Mbps; Flow 3 is ensured with 0.25 Mbps.

Weighted queue: $w_1=2, w_2=1, w_3=1$

Efficiency:

When flow 3 has nothing to transmit, but flow 1 and flow 2 have many packets to send

Flow 1: 2/3 Mbps

Flow 2: 1/3 Mbps

Marking and Policing

Policing mechanisms

goal: to limit traffic to not exceed declared parameters (the rate at which a class or flow is allowed to inject packets into the network)

› *Three important policing criteria (differing on the time scale):*

› *1. (long term) average rate:* how many pkts can be sent per unit of time (in the long run)

- e.g., 6000 packets per min

- 2.*peak rate:* limit the number of packets can be sent over a relatively shorter period of time, e.g., 6000 pkts per minute (ppm) in average but 3000 packets per 5 second peak rate max.

- *3. (max.) burst size:* max number of pkts sent "instantaneously" into the networks, e.g., 1500 packets.

Policing mechanisms: implementation

token bucket: limit input to specified *burst size* and *average rate* (useful to police the flow)

› bucket can hold b tokens

› a packet must remove a token from bucket to be transmitted into the network

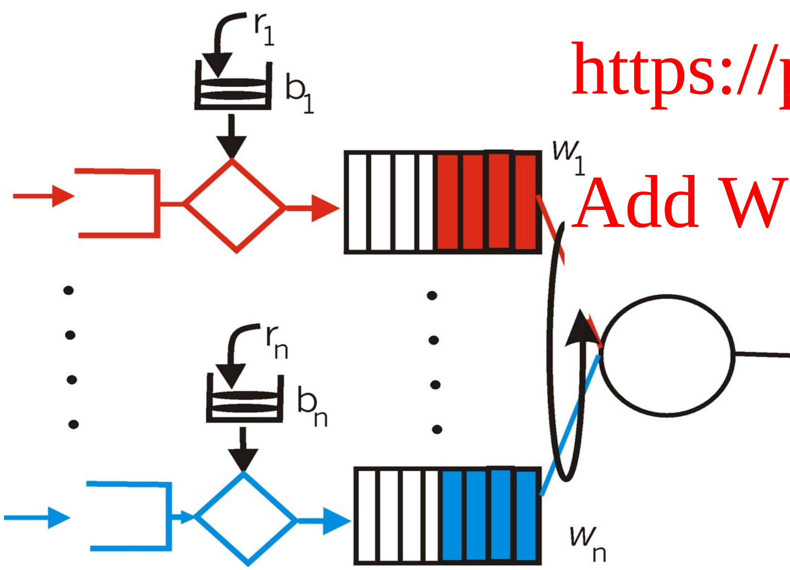
› tokens generated at rate *r token/sec* unless bucket full (token ignored)

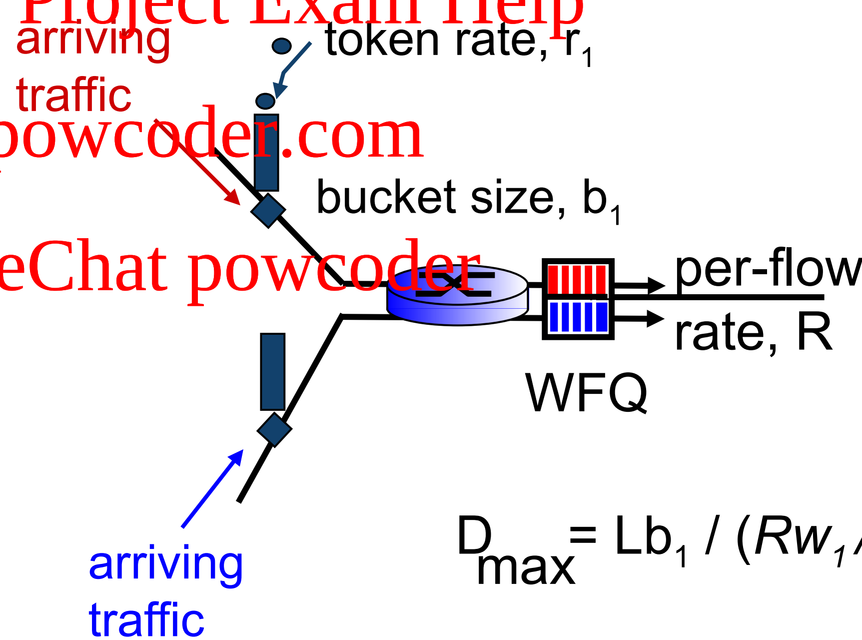
› *over interval of length t: number of packets admitted less than or equal to (rt + b)*

› *Token-generation rate r limits the rate at which packets enter the network*
t->0, b packets *t->∞, (rt+b)/t = r packets/second*

Policing and QoS guarantees

- Combining token bucket and WFQ to provide guaranteed upper bound on delay, i.e., *QoS guarantee!*





$$D_{max} = Lb_1 / (Rw_1 / (\sum w_j))$$

- Packets arrive while the bucket is full (b_1). The last packet has a maximum delay of D_{max}. L packet size.

Differentiated services in reality

- want "qualitative" service classes
  - relative service distinction: Platinum (VIP), Gold, Silver
- *scalability:* simple functions in network core, relatively complex functions at edge routers (or hosts)

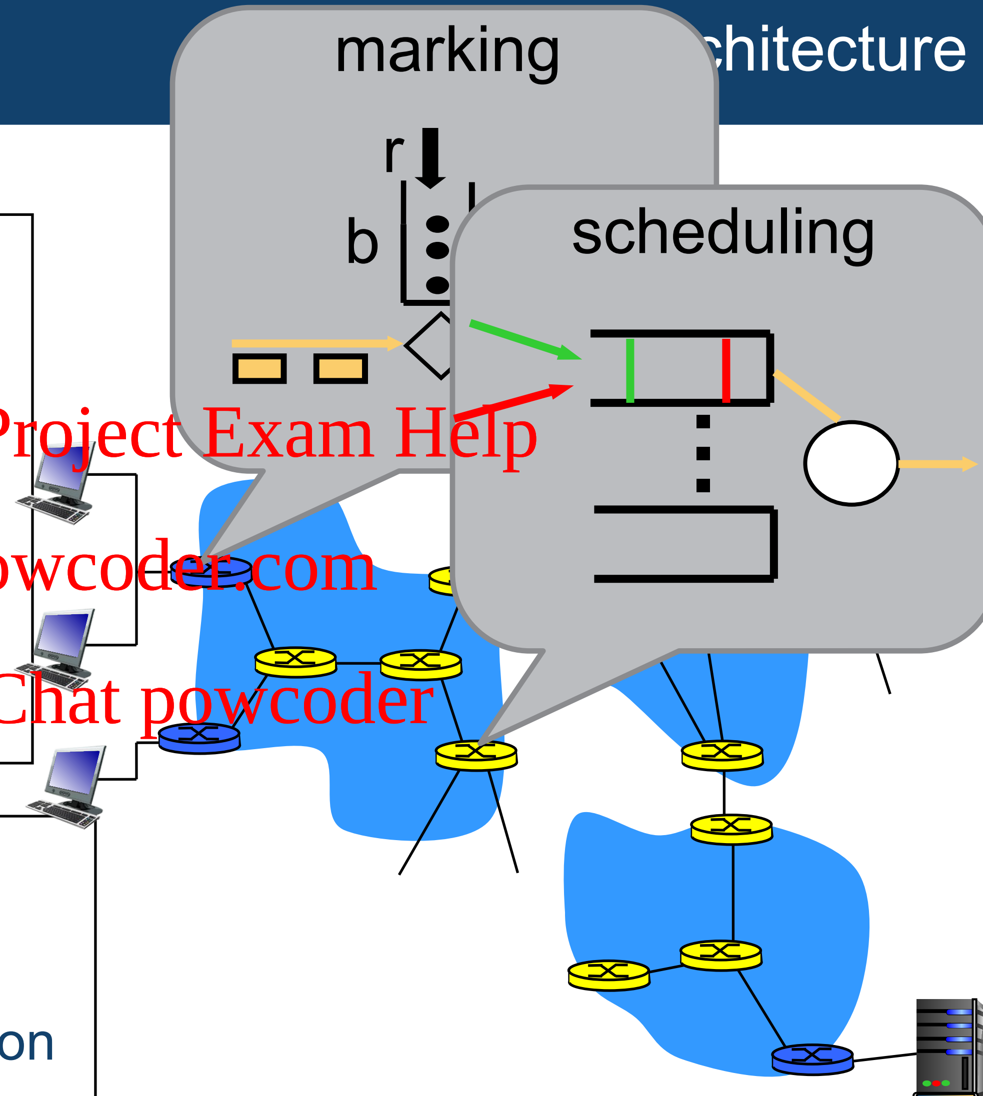
chitecture

edge router:

- per-flow traffic management
- marks packets as differently
  - E.g. Alice' traffic : high
  - Bob's traffic: high
  - Chris's traffic: low

core router:

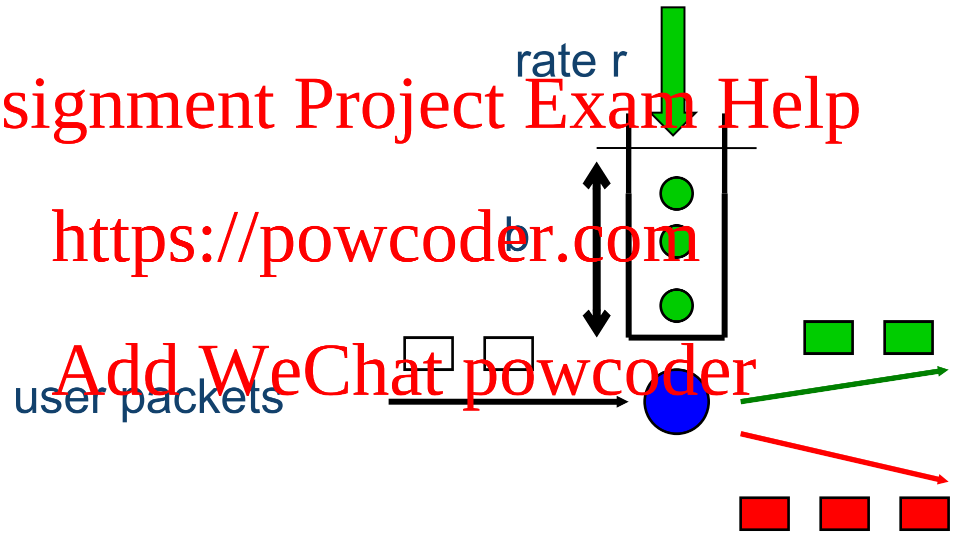
- per class traffic management
- buffering and scheduling based on marking at edge
- Red packets vs green packets



Edge-router packet marking

› profile: pre-negotiated rate r, burst (bucket) size b

› packet marking at edge based on per-flow profile



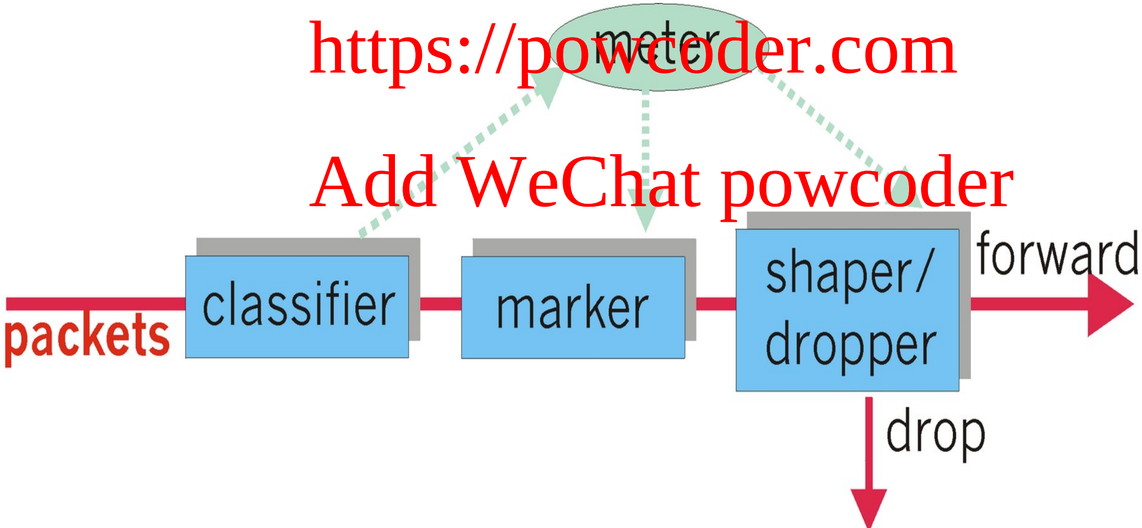
Example:

- class-based marking: packets of different classes marked differently
- intra-class marking: conforming portion of flow marked differently compared with non-conforming one
  - Bob agrees to transmit at 1Mbps, but he is transmitting at 2Mbps
  - Half of them (conforming) are marked green.
  - Others (non-conforming) are marked red (lower priority) or dropped.

- Example
- green > yellow >red .
  - 2Mbps link, Bob, telephone traffic, declare 1 Mbps
    - Green if conforming, red if not conforming
  - Chris, web browsing traffic
    - Yellow
  - Priority queue in the core network
- Bob can guarantee 1Mbps data rate
- If Bob transmits >1Mbps
  - If Chris transmits at 1Mbps, all red will be dropped. Bob gets 1Mbps
  - If Chris transmits at <1Mbps, some red will still get through.

Classification, conditioning

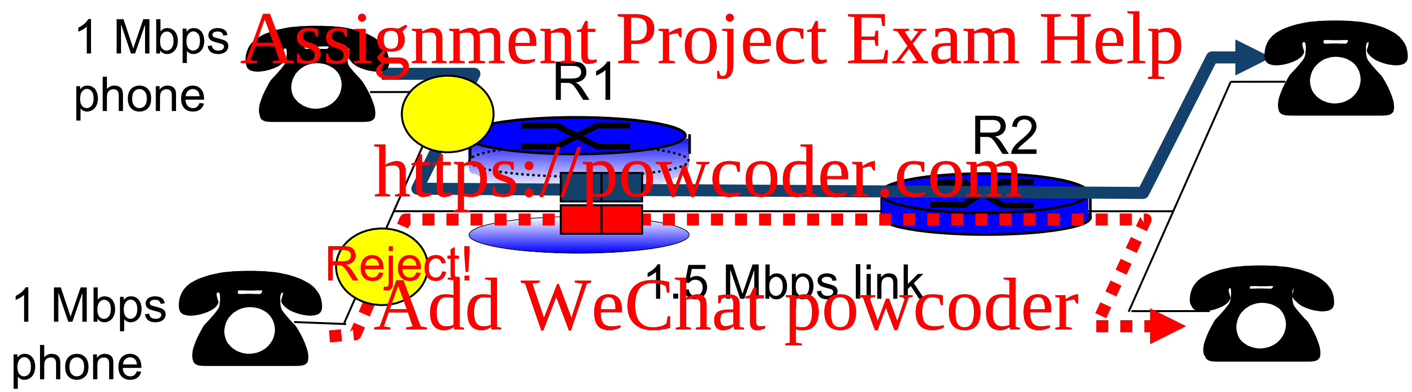
- user declares traffic profile (e.g., rate, burst size)
- traffic metered, shaped if non-conforming



- the meter compares the incoming flow to the negociated traffic profile. Network administrator can decide whether to remark, forward, delay, or drop a non- conforming packet

Per-connection QoS guarantees

› *basic fact of life:* cannot support traffic demands beyond link capacity

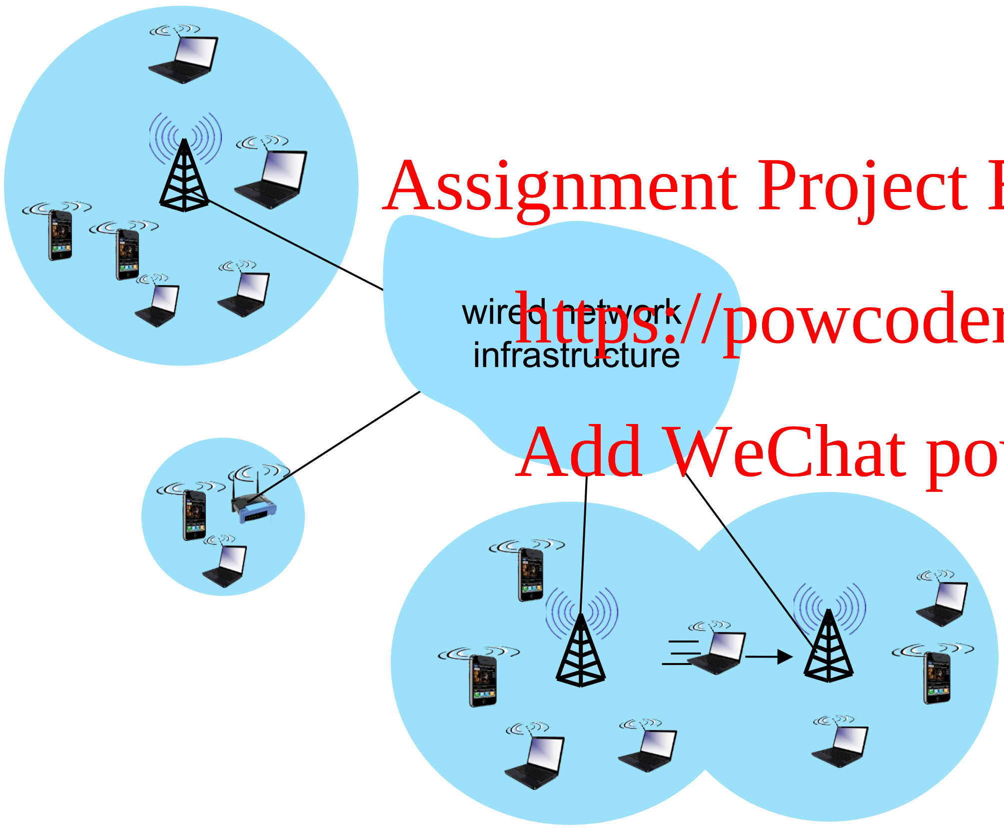


Principle 4

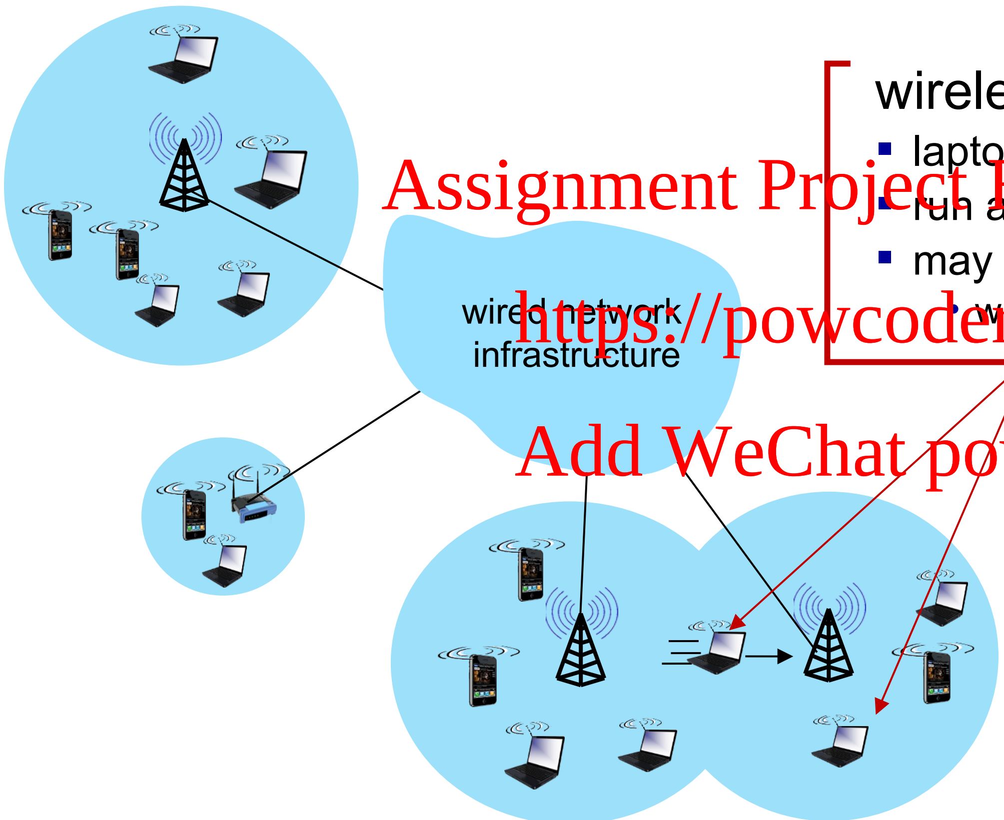
call admission: flow declares its needs, network may block call (e.g., busy signal) if it cannot meet needs

Wireless and mobile networks

Elements of a wireless network



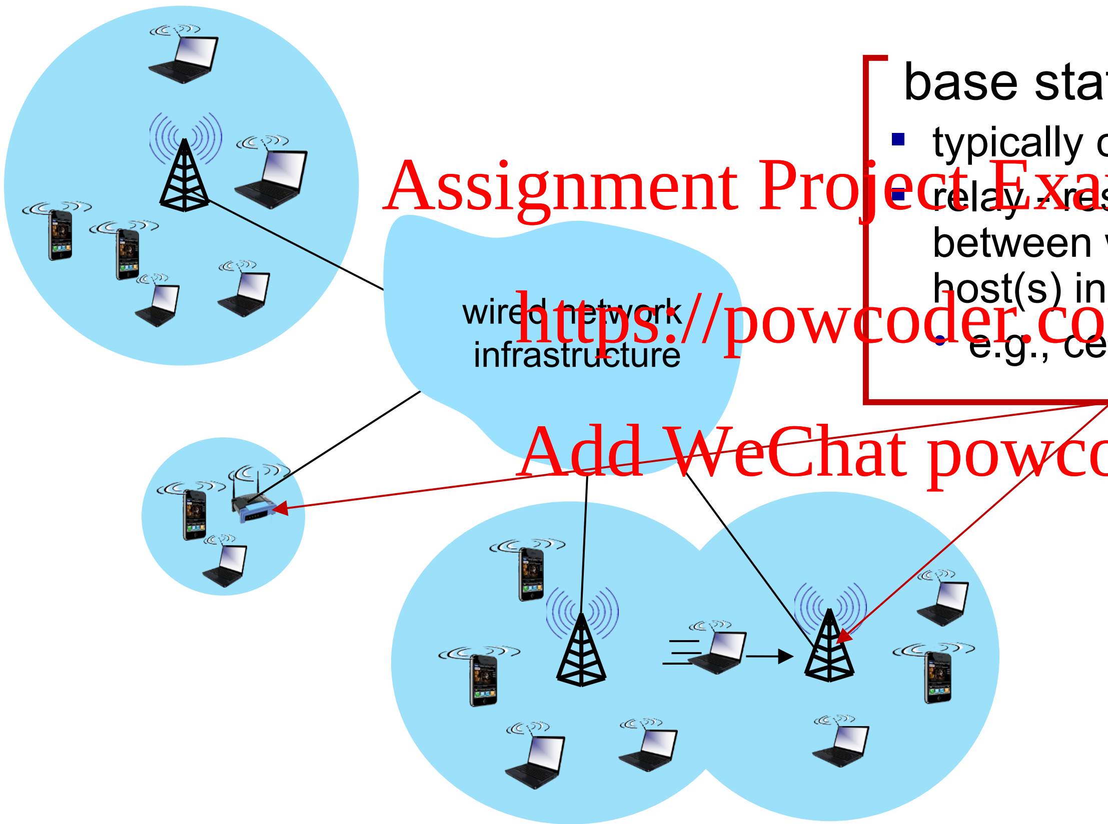
Elements of a wireless network



wireless hosts

- laptop, smartphone, IoT
- run applications
- may be stationary (non-mobile) or mobile
  - wireless does *not* always mean mobility!

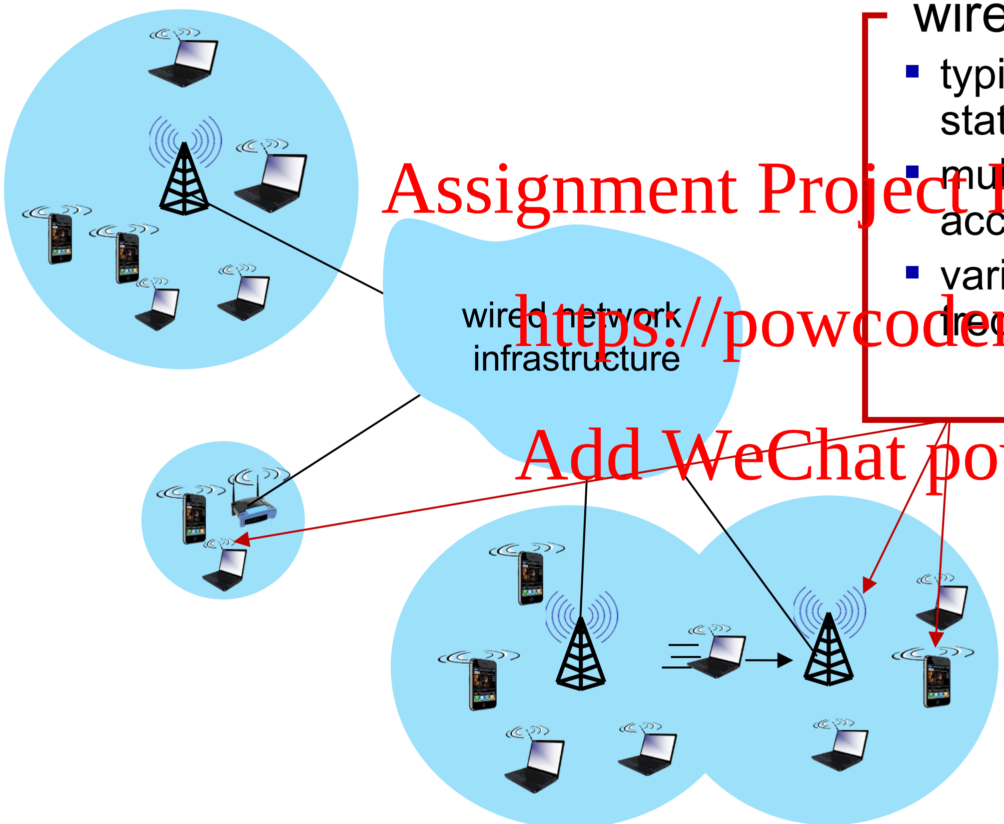
Elements of a wireless network



base station

- typically connected to wired network
- relay - responsible for sending packets between wired network and wireless host(s) in its "area"
  - e.g., cell towers, 802.11 access points

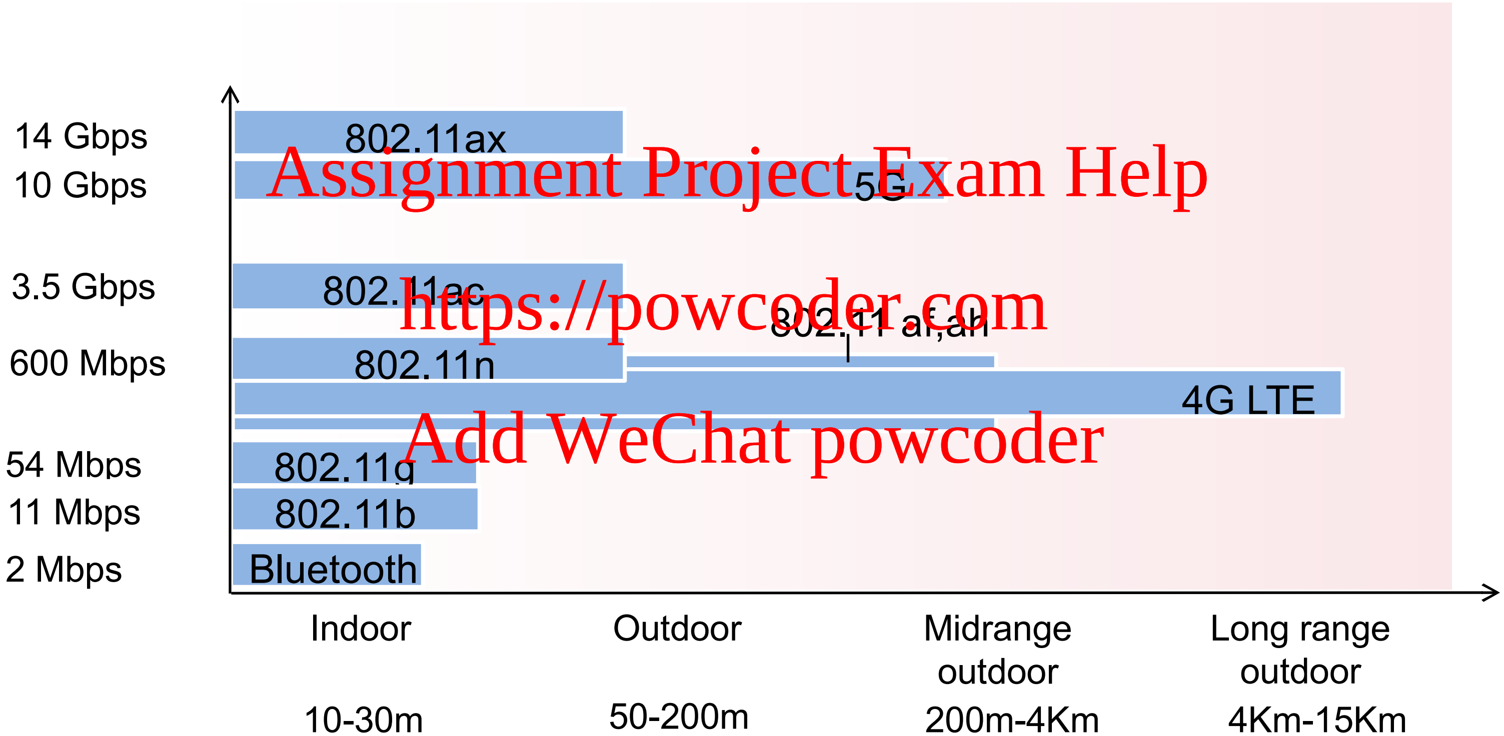
Elements of a wireless network



wireless link

- typically used to connect mobile(s) to base station, also used as backbone link
- multiple access protocol coordinates link access
- various transmission rates and distances, frequency bands

Characteristics of selected wireless links

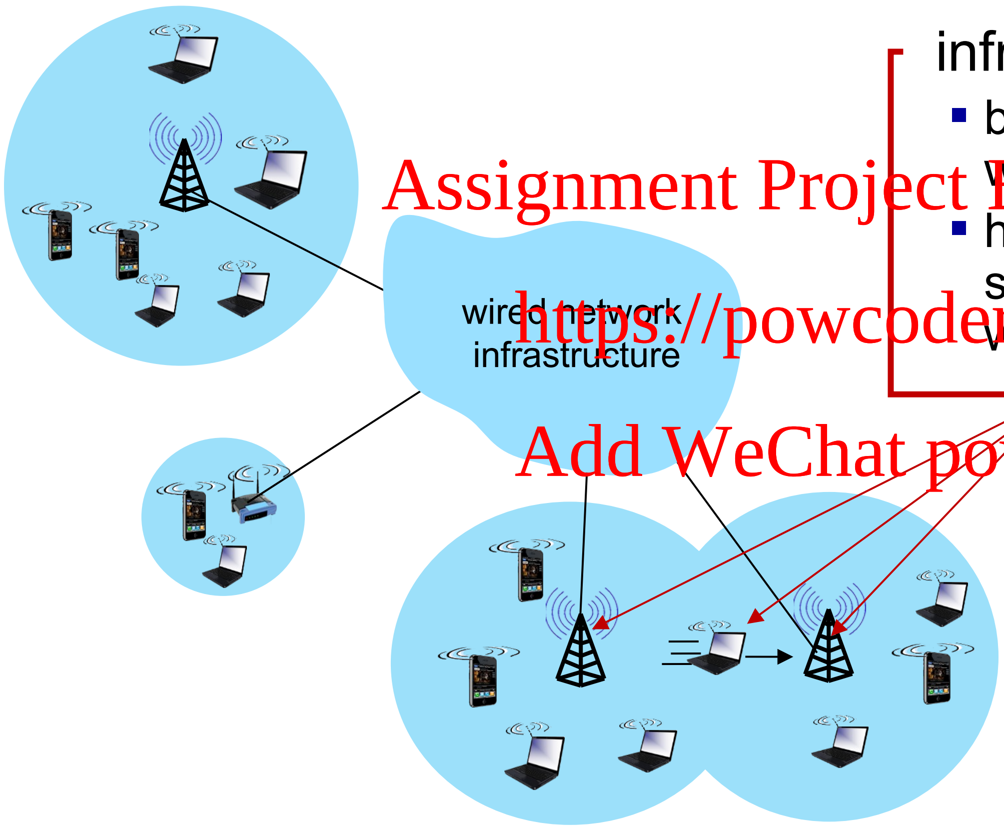


IEEE 802.11 WiFi

IEEE 802.11 standard	Year	Max data rate	Range	Frequency
802.11b	1999	11 Mbps	30m	2.4 Ghz
802.11a	1999	54 Mbps	30m	5 Ghz
802.11g	2003	54 Mbps	30m	2.4 Ghz
802.11n (WiFi 4)	2009	600 Mbps	70m	2.4, 5 Ghz
802.11ac (WiFi 5)	2013	3.47Gpbs	70m	5 Ghz
802.11ax (WiFi 6)	2020	14 Gbps	70m	2.4, 5 Ghz
802.11af	2014	35 – 560 Mbps	1 Km	unused TV bands (54-790 MHz)
802.11ah	2017	347Mbps	1 Km	900 Mhz

- all use CSMA/CA for multiple access, and have base-station and ad-hoc network versions

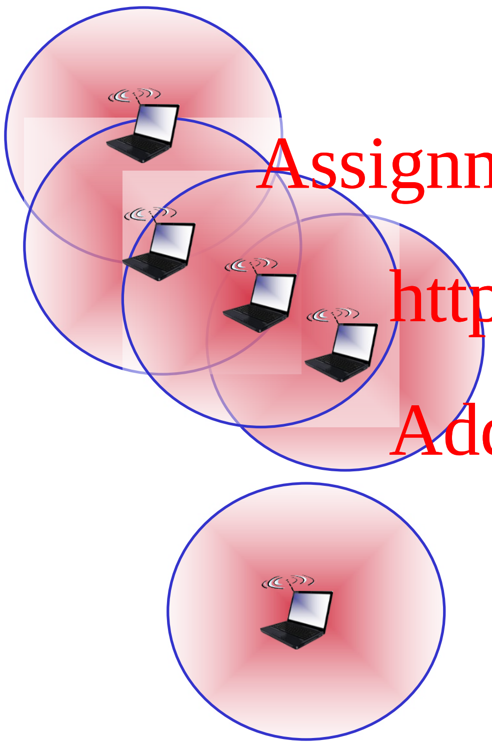
Elements of a wireless network



infrastructure mode

- base station connects mobiles into wired network
- handoff: mobile changes base station providing connection into wired network

Elements of a wireless network



ad hoc mode

- no base stations
- nodes can only transmit to other nodes within link coverage
- nodes organize themselves into a network: route among themselves

THE UNIVERSITY OF SYDNEY

Wireless network taxonomy

	single hop	multiple hops
infrastructure (e.g., APs)	host connects to base station (WiFi, cellular) which connects to larger Internet	host may have to relay through several wireless nodes to connect to larger Internet: *mesh net*
no infrastructure	no base station, no connection to larger Internet (Bluetooth, ad hoc nets)	no base station, no connection to larger Internet. May have to relay to reach other a given wireless node MANET, VANET

Wireless link characteristics

Wireless Link Characteristics (1)

important differences from wired link ….

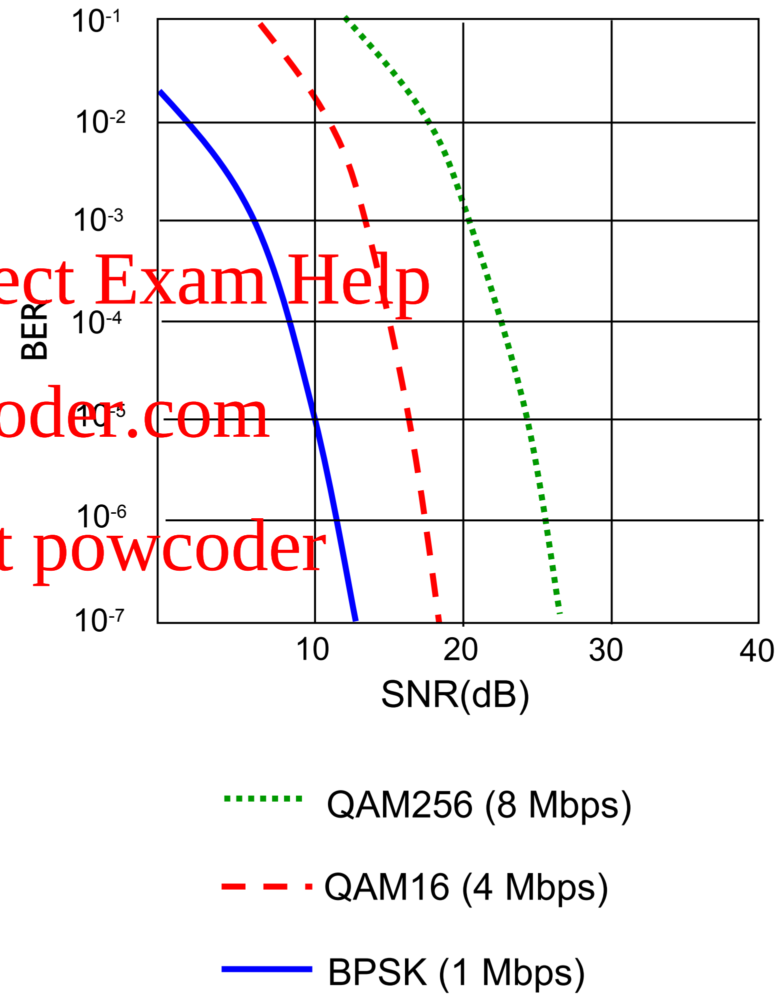
- *decreased signal strength:* radio signal attenuates as it propagates through matter (path loss)
- *interference from other sources:* standardized wireless network frequencies (e.g., 2.4 GHz) shared by other devices (e.g., phone);
- *multipath propagation:* radio signal reflects off objects ground, arriving at destination at slightly different times

…. make communication across (even a point to point) wireless link much more "difficult"

dB decibel

- logarithmic unit used to express the ratio of two (power) values
- $10*\log_{10}(P_S/P_N)$
- $P_S/P_N=10$ 10 dB
- $P_S/P_N=100$ 20 dB
- $P_S/P_N=1000$ 30 dB
- $P_S/P_N=10000$ 40 dB
- …

Wireless Link Characteristics (2)



- SNR: signal-to-noise ratio
  - larger SNR – easier to extract signal from noise (a "good thing")
  - BER: bit error rate
- *SNR versus BER tradeoffs:*
  - *given physical layer modulation:* increase power -> increase SNR -> decrease BER
  - *Different physical layer modulation:*

Quadrature amplitude modulation

Binary Phase-shift keying

Higher data rate -> Higher BER

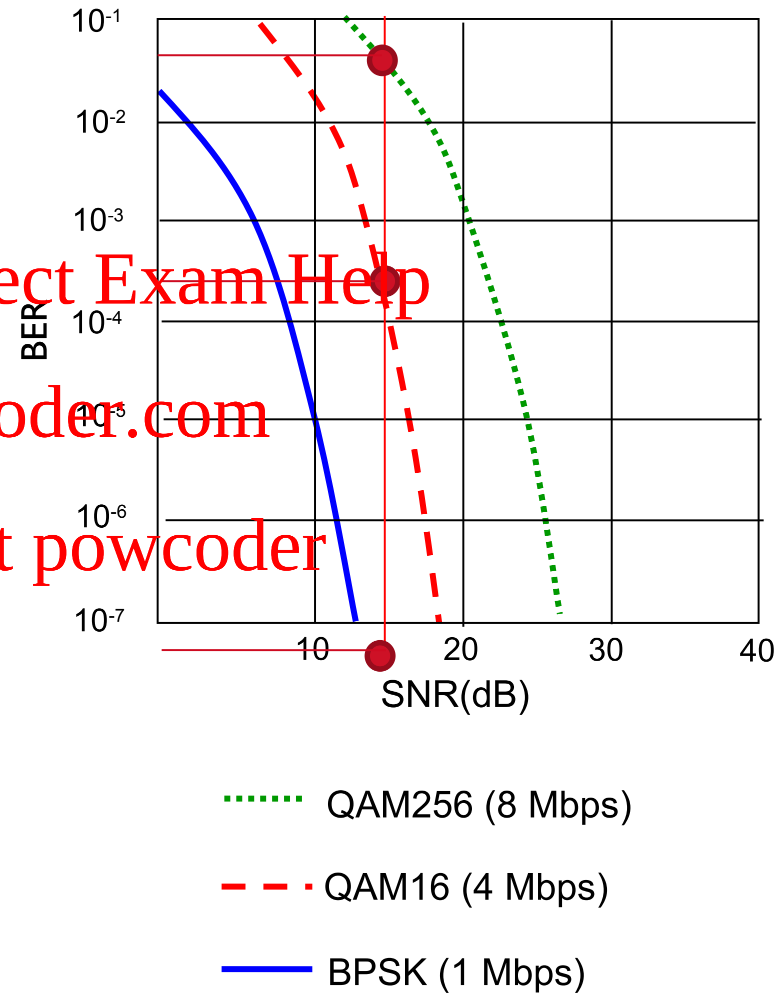
Wireless Link Characteristics (2)

› SNR: signal-to-noise ratio

- larger SNR – easier to extract signal from noise (a "good thing")
- BER: bit error rate

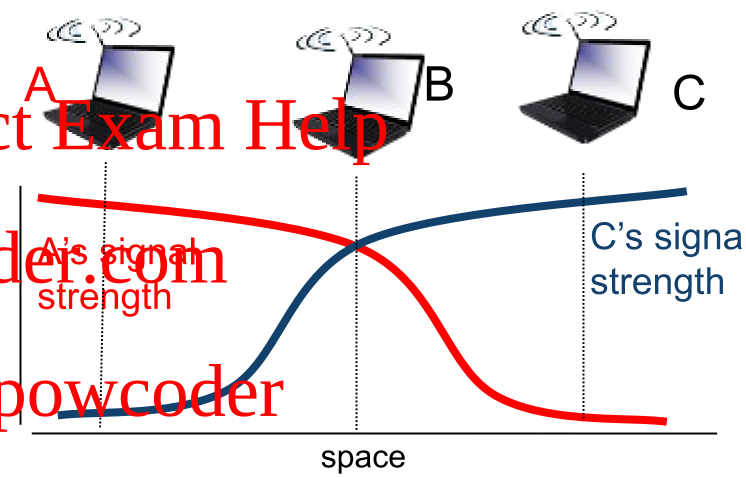
› *SNR versus BER tradeoffs*

- *given SNR, BER requirement:* choose modulation to achieve highest throughput
  - 15 dB, require 10^{-3} BER
  - Which modulation?
  - QAM16



Wireless network characteristics (3)

Multiple wireless senders and receivers create additional problems (beyond multiple access):



Hidden terminal problem

- B, A hear each other
- B, C hear each other
- A, C can not hear each other means A, C unaware of their interference at B

Signal attenuation:

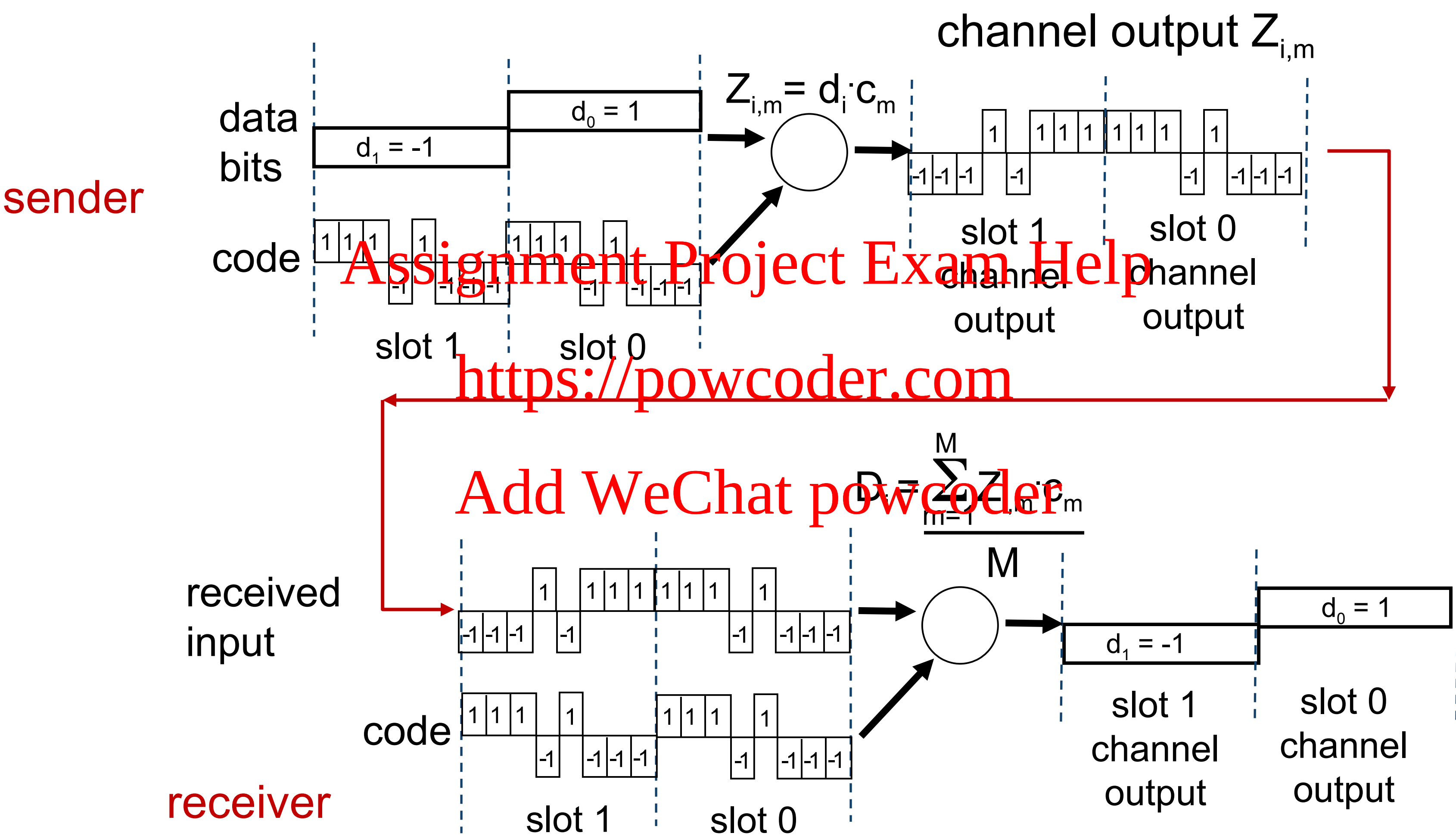
- B, A hear each other
- B, C hear each other
- A, C can not hear each other interfering at B

CDMA

Code Division Multiple Access (CDMA)

- unique "code" (chipping sequence) assigned to each user;
- all users share same frequency, but each user has own "chipping" sequence (i.e., code) to encode data
  - length of sequence: M
  - allows multiple users to "coexist" and transmit simultaneously with minimal interference (if codes are "orthogonal")
  - orthogonal:
    - inner product of $c_{i,1}\, c_{i,2} \ldots c_{i,M}$ and $c_{j,1}\, c_{j,2} \ldots c_{j,M}$ is $\sum_m c_{i,m} \cdot c_{j,m}$
    - inner product(user i's chipping sequence, user j's chipping sequence) =0
    - inner product(user i's chipping sequence, user i's chipping sequence) =M
- *encoded signal* = (original data) X (chipping sequence)
- *decoding:* inner-product of encoded signal and chipping sequence

CDMA encode/decode



User i receives user i's signals

sender	d_i	1	-1
	c_m	-1 -1 -1 1 -1 1 1 1	-1 -1 -1 1 -1 1 1 1
	$Z_{i,m}$	-1 -1 -1 1 -1 1 1 1	1 1 1 -1 1 -1 -1 -1
receiver		-1 -1 -1 1 -1 1 1 1	1 1 1 -1 1 -1 -1 -1
Inner product	$D_i = \frac{\sum_{m=1}^{M} Z_{i,m} \cdot c_m}{M}$	-1 -1 -1 1 -1 1 1 1	-1 -1 -1 1 -1 1 1 1
		1+1+1+1+1+1+1+1	
	D_i	8/8=1	-8/8=-1

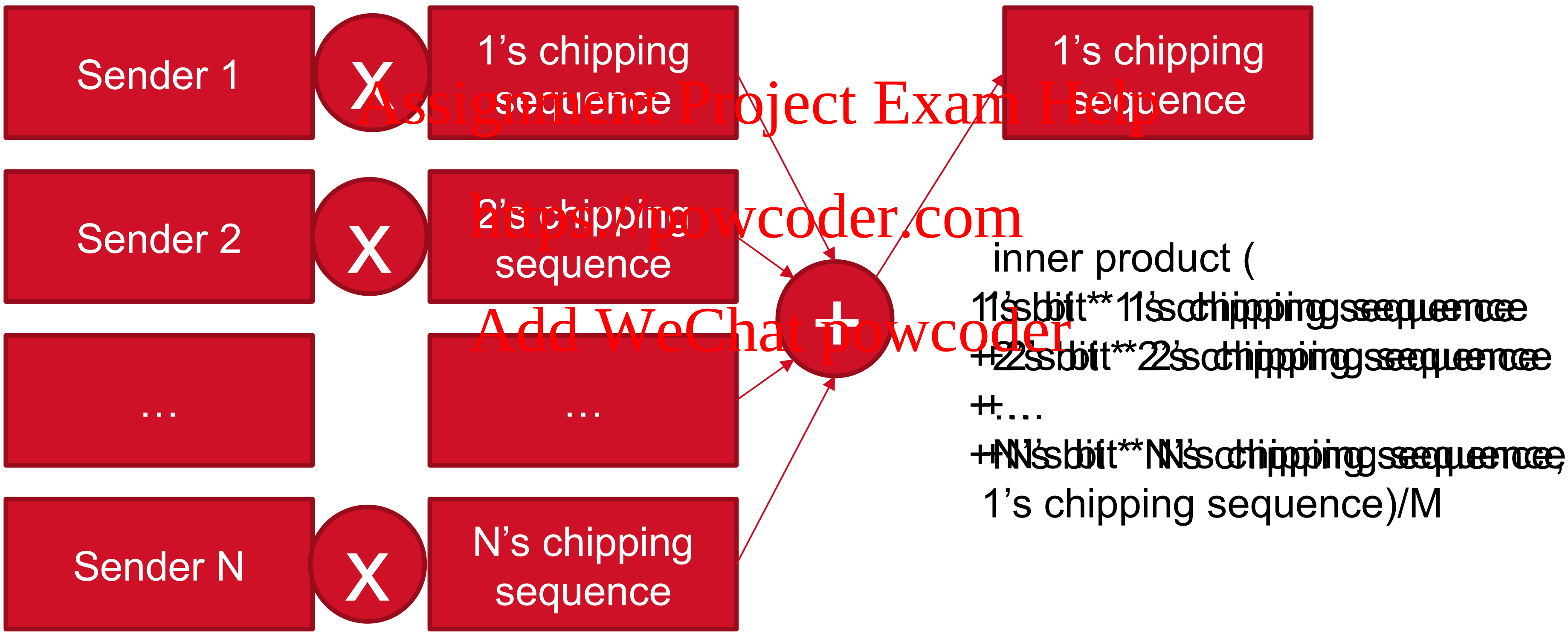
uses its chipping sequence to send and to receive: receive the correct bits

User 2 receives user 1's signals

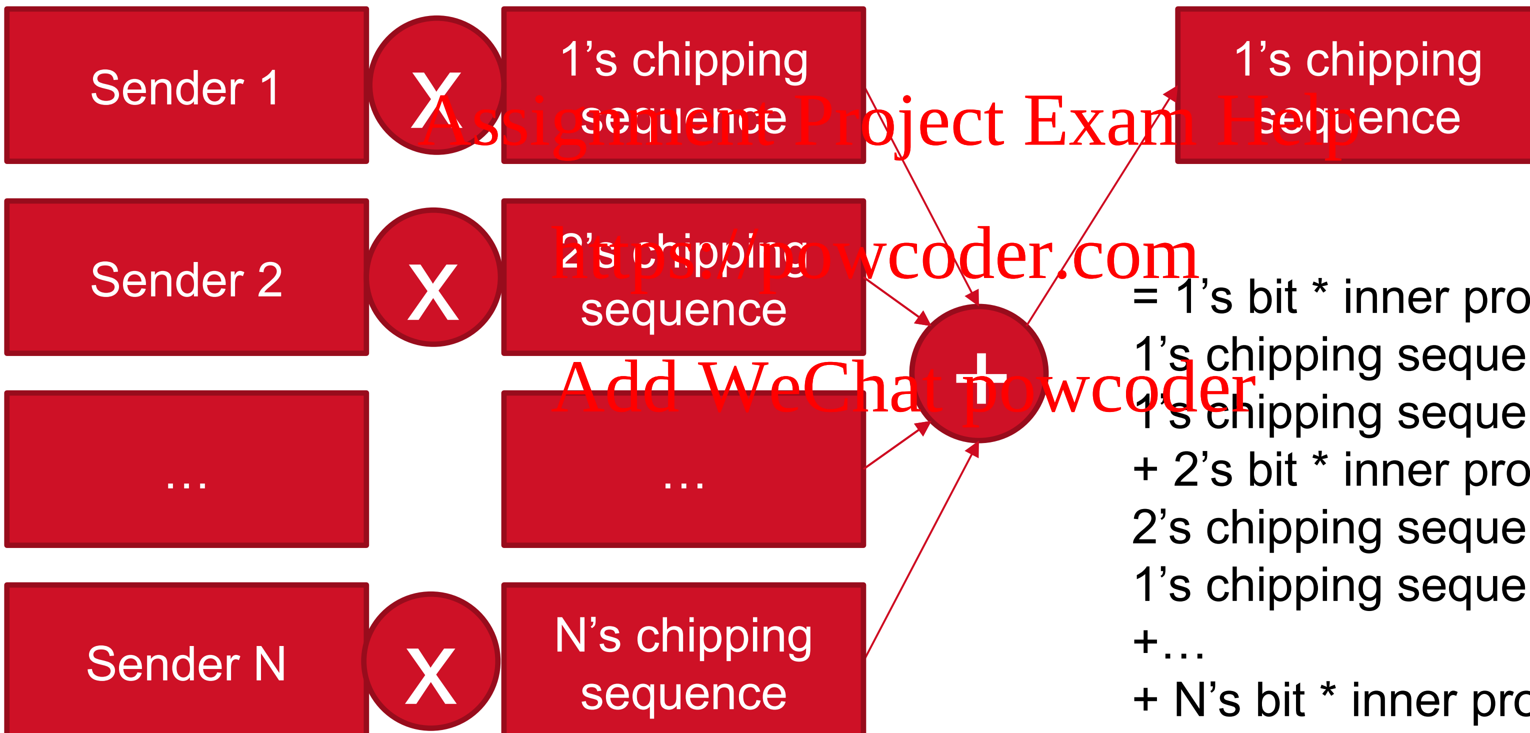
sender 1's bits	d_i	1	-1
	c_m	-1 -1 -1 1 -1 1 1 1	-1 -1 -1 1 -1 1 1 1
	$Z_{i,m}$	-1 -1 -1 1 -1 1 1 1	1 1 1 -1 1 -1 -1 -1
receiver 2	$D_i = \frac{\sum_{m=1}^{M} Z_{i,m} \cdot c'_m}{M}$	-1 -1 -1 1 -1 1 1 1	1 1 1 -1 1 -1 -1 -1
Inner product		1 1 -1 1 1 1 -1 1	1 1 -1 1 1 1 -1 1
		-1-1+1+1-1+1-1+1=0!	
	D_i	0/8=0	0/8=0

Use 1's chipping sequence to send and use 2's chipping sequence to receive: receive nothing!

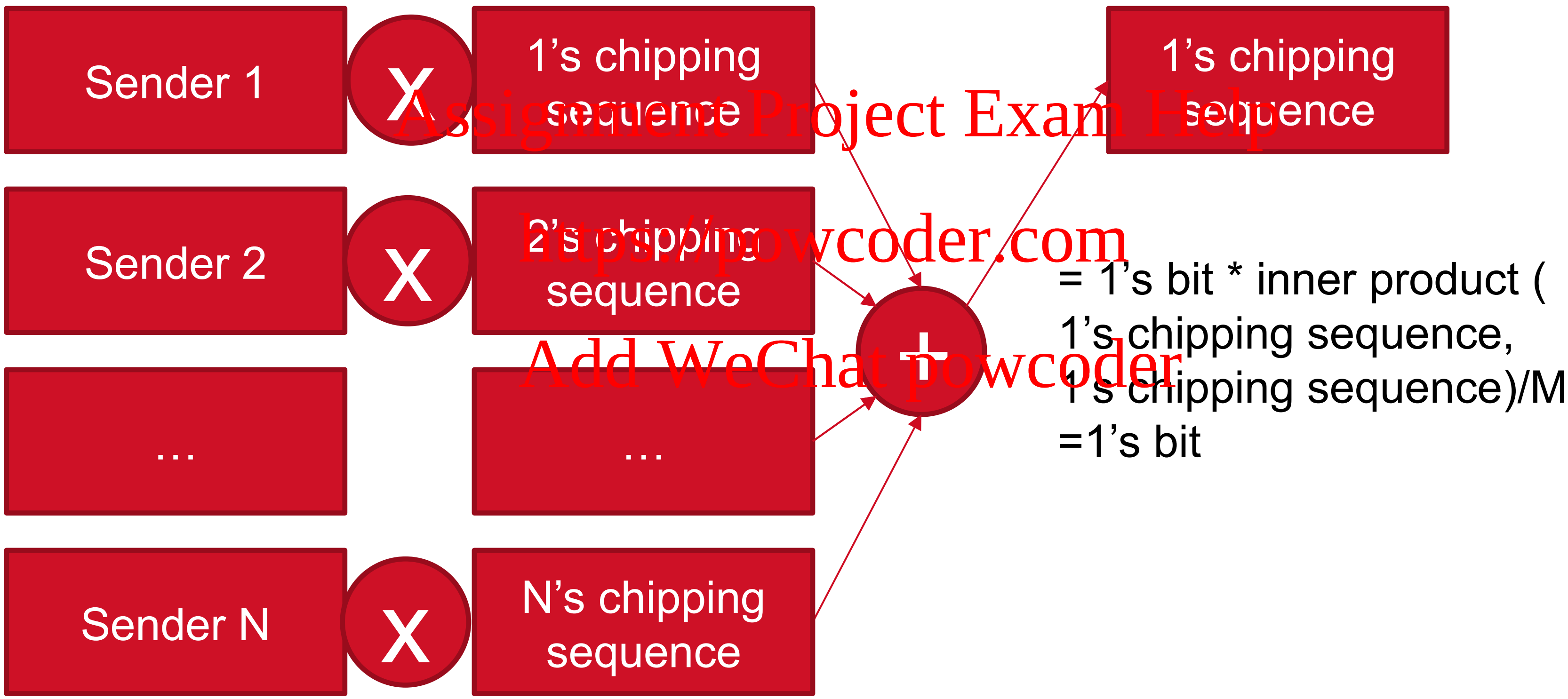
Reason: 1's chipping sequence is orthogonal to 2's chipping sequence.



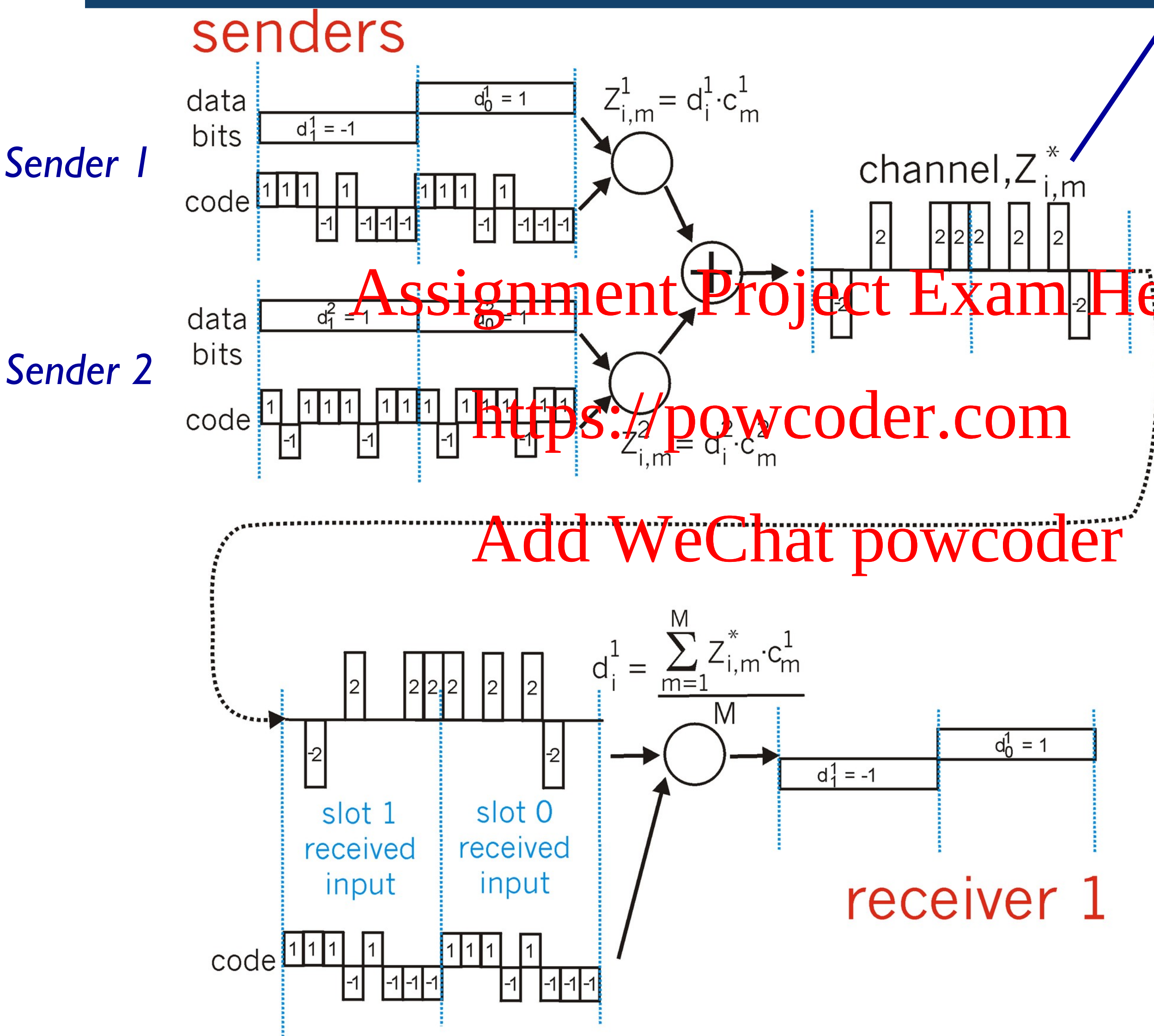
Sender 1
X
1's chipping sequence
Sender 2
X
2's chipping sequence
…
…
Sender N
X
N's chipping sequence
+
1's chipping sequence
inner product (
1's bit * 1's chipping sequence
+2's bit * 2's chipping sequence
+…
+N's bit * N's chipping sequence,
1's chipping sequence)/M



= 1's bit * inner product (1's chipping sequence, 1's chipping sequence)/M
+ 2's bit * inner product (2's chipping sequence, 1's chipping sequence)/M 0
+…
+ N's bit * inner product (N's chipping sequence, 1's chipping sequence)/M 0


Sender 1
X
1's chipping sequence
Sender 2
X
2's chipping sequence
…
…
Sender N
X
N's chipping sequence
+
1's chipping sequence
= 1's bit * inner product (
1's chipping sequence,
1's chipping sequence)/M
=1's bit

CDMA: two-sender interference



channel sums together transmissions by sender 1 and 2

using same code as sender 1, receiver recovers sender 1's original data from summed channel data!